Thomas Quinn
Department of Astronomy
Box 351580
University of Washington
Seattle, WA 98195-1580
phone: 206-685-9009
FAX: 206-685-0403
email: `trq@astro.washington.edu`

September 29, 2009

Dr. Joseph H. Bredekamp
Senior Science Program Executive/Information Systems
Heliophysics Division
Science Mission Directorate
NASA Headquarters
Washington, DC 20546-0001

Dear Dr. Bredekamp,

RE: Second year report and request for renewal of NASA AISRP grant NNX08AD19G: Visualization of Terascale Datasets with Impostors

Principle Investigator: Thomas Quinn

Co-Investigator: Laxmikant V. Kale

Co-Investigator: Orion Sky Lawlor

During the past year the investigators continued their collaboration on developing innovative techniques for visualization and analysis of large datasets. This work is building on previous work that the participants have done together. Specific implementations are being built on top of our existing simulation, analysis and visualization frameworks, partially funded by a previous AISRP grant. Planning and coordination of this effort was expedited by an in-person meetings among the investigators at the Charm++ workshop at University of Illinois. This workshop also provided a venue for communicating our results. Our results were also presented at an invited talk at the NSF Teragrid conference in Washington D.C. Telecons were held on a bi-weekly schedule throughout the rest of the year to keep the groups at all three institutions abreast of progress.

At the University of Alaska Fairbanks, Lawlor and his student integrated 3D voxel grids representing particle data, or "volume impostors", into our particle visualization system SALSA. First, we added volume impostor rendering to the visualization server, and soon realized we had to tune the rendering server to perform acceptably with the additional rendering load. Next, we extended our client-server network interface, liveViz, to assemble and communicate 3D voxel grids, currently stored as a vertical stack of 2D image slices. To render 3D volume impostors, we adjusted the Java client to access the graphics hardware directly via JOGL. Today, volume impostors are stored as 3D textures directly on the client's graphics card, rendered slice-by-slice into an offscreen buffer and then run through the current color table using a pixel shader. Currently, we are working to make the client perform better over slow links, and be easier to use to extract real scientific insight from our large datasets.

Quinn and his graduate students at the University of Washington have been testing these rendering schemes for scientific usability with datasets we are currently using for cosmology and galactic structure. The UW group has access to machines with a variety of 3D hardware capabilities, and we have been using these to enhance the portability of the rendering schemes. In addition his group has been adding scientific functionality to the SALSA visualization framework in order to turn it into a useful integrated visualization and analysis system.

In the Illinois group led by Prof. Kale, most of the work conducted during this period focused on improvements to the ChaNGa cosmological simulator, which generates the large datasets that we wish to visualize, and which incorporates an on-the-fly parallel visualization module. Our improvements comprised both new functionalities in the code and optimizations to the existing version that is distributed to the community. These optimizations are beneficial for simulation, visualization and analysis.

One of the enhancements made in ChaNGa was a refactoring of the SSE gravity computation, making it substantially more compact and speeding the sequential execution up by 12% when using the float SSE version. SPH performance was also optimized by replacing the STL priority queue with an array maintained as a heap, improving overall performance by 15%. Finally, a refactoring of the SPH code was done, using inheritance to get rid of code duplication.

Improving the performance of the parallel version of ChaNGa was another active area of work. This has included work on parallel sorting with various schemes of merging the communicated data and different approaches to determining splitters. An internally developed parallel sorting benchmark was used for testing different ways of doing the histogramming algorithm and varying various parameters to determine what works best. Meanwhile, there were several experiments with the load balancing framework and load distribution strategies in ChaNGa. Preliminary results have been encouraging: it was possible to improve the code performance on a 5 million particle dataset running on 1024 processors of Intrepid (IBM-BG/P) by 12 percent. Other enhancements have been added to the code, such as a more flexible scheme for the domain decomposition, which allows better distribution of work to processors at runtime.

In order to leverage the capabilities of current advanced GPUs, the Illinois group has produced an API for GPU execution that allows the user to offload pieces of work (work requests) to the GPU, with the system ensuring overlap between execution of a work request and data transfer from and to the GPU when multiple requests are queued up. Using this API, the Illinois group created a CUDA port of the application and is currently experimenting to gauge its performance. Initial results have shown the CUDA version to be more than twice as fast as the original one on a small, 3 million particle dataset. With more effort expended on tuning the application and the use of larger datasets, more appreciable gains in performance are expected.

In the coming year we will both incorporate some of the outcomes of this years development into our production codes, and continue to develop more advanced visualization techniques. The JPEG image compression will be integrated into our SALSA parallel visualization framework. Work will also continue on making the on-the-fly visualization of large simulations more interactive, and therefore, more useful. This involves both improving the performance of the simulation codes, and experimenting with techniques that can overcome latency. We will also be experimenting with efficient compositing of 3D "volume impostors" on the parallel machine and transmitting them across the network. Since this is the final year of this project, we will be preparing code releases to the community of the visualization and simulation frameworks.

In light of our progress, I request that this grant be renewed for a second year in the amount of $244,181 for the period 11/30/09 to 11/29/10. The budget for spending this money is as given on the third year budget form submitted with the original proposal, with further details given in the subcontracts Appendix to that proposal.

Sincerely,

Thomas Quinn